



SparkCognition DeepNLP™ Installation Guide

**A SparkCognition™ Education Document
Q1 - 2019**

This document contains copyrighted and proprietary information of SparkCognition and is protected by United States copyright laws and international treaty provisions. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under such laws or with the prior written permission of SparkCognition Inc.

SparkCognition[™], the sparkcognition logo, Darwin[™], DeepArmor[®], DeepNLP[™], MindFabric[®], SparkSecure[®] and SparkPredict[™], are trademarks of SparkCognition, Inc. and/or its affiliates and may not be used without written permission. All other trademarks are the property of their respective owners.

©SparkCognition, Inc. 2017-2019. All rights reserved.

SparkCognition DeepNLP™ Installation Guide

Contents

About this guide	1
Requirements	2
Bare-metal Installation Requirements	2
Virtual Machine Installation Requirements	2
Installation & Configuration Process	3
Extraction	3
Single-Instance Installation	3
Load Image and Start DeepNLP	3
Multi-Instance Installation	4
Reference	4
Hardware Requirements	4
Operating System/Software Requirements	4
Services	4
Multi-host Services	4
Application Access Requirements	4

About this guide

This guide describes the Sparkcognition DeepNLP™ Installation process. This guide is intended for data scientists, software engineers and analysts who want to interact with the DeepNLP system. Note that throughout this document, some long key and token values are truncated - indicated by ellipses (...) - or broken across lines, indicated with a backslash (\).

Requirements

The general requirements for DeepNLP include:

- Two DeepNLP files - supplied by SparkCognition, either attached as part of the SparkCognition welcome email or available from SparkCognition DeepNLP support:
 - `Deepnlp-docker-bundle.tar` - tar file containing docker-compose configuration files and other installation necessities
 - `Deepnlp-images.tar` - tar file containing Docker images
 - Operating system - Either *CentOS 7* or *RedHat 7*
 - Appropriate machine - DeepNLP can be installed on virtual or physical machines running supported operating systems
- Note:** DeepNLP is not supported if the requirements are not met or exceeded

Bare-metal Installation Requirements

- *CentOS 7* or *RedHat 7*
 - Hardware requirements include:
 - Minimum: 4 modern CPU Cores, 32GB memory, 30GB Storage
 - Recommended: 16 modern CPU Cores, 64GB memory, 60GB Storage
 - A network connection to the outside world must be present during installation to install required system and Python dependencies
- Note:** The outbound network connection must not be restricted by a proxy
- The user installing DeepNLP requires *ssh access* to the DeepNLP installation target machines and access to a control server to run the installer from
 - The user installing DeepNLP requires *sudo* privileges to the machines on which DeepNLP is to be installed

Virtual Machine Installation Requirements

Virtual machines (VM) installed with DeepNLP must meet all above requirements, including those for memory and processor.

DeepNLP requires VM hosts that are compatible with *VMWare Workstation™ 15* and support *VMWare Workstation 15* images.

Contact SparkCognition for additional information concerning support for other VMWare® hosting products or versions of VMWare.

Installation & Configuration Process

The following sections describe the preparation for installation and configuration of DeepNLP.

Extraction

To extract the `deepnlp.tgz` file, issue the following commands as a user with `sudo` authority:

```
sudo tar zxvf deepnlp.tgz
sudo cd deepnlp
```

Single-Instance Installation

To install a single instance of DeepNLP, perform the following:

1. Copy the `Deepnlp-docker-bundle.tar` file to a folder (your choice)
2. Extract the file
3. Run the following commands as a user with `sudo` authority:

```
cd deepnlp-docker-bundle/offline
sudo ./offline_install install <USERNAME>
```

Where `USERNAME` is a username of your choice - your own username on the system, for example.

This installs `docker` and `docker-compose`. It places the `deepnlp-docker` folder within the home directory of `<USERNAME>`. If `<USERNAME>` is not provided, the default user name `deepnlp` is used. The installation adds `<USERNAME>` to the `docker` group to enable `docker` to run without the need to use `sudo`.

Load Image and Start DeepNLP

Perform the following to load the new docker images and start DeepNLP:

1. Login as `<USERNAME>`
2. Issue the following commands as a user with `sudo` authority :

```
docker load --input=Deepnlp-images.tar
sudo su <USERNAME>
cd /home/ <USERNAME> /deepnlp-docker
./deepnlp.sh start # make sure you are logged in as USERNAME
```

Multi-Instance Installation

DeepNLP supports multi-instance installation. Because that procedure is beyond the scope of this document, contact SparkCognition DeepNLP support for assistance to deploy multiple instances.

Reference

Hardware Requirements

- Minimum: 4 modern CPU Cores, 32GB memory, 30GB Storage
- Recommended: 16 modern CPU Cores, 64GB memory, 60GB Storage

Operating System/Software Requirements

- *CentOS 7 or RedHat 7*

Services

The two main DeepNLP components are the web frontend and the scheduler. These components expect the following services to be running:

- *Postgres 9.5*, or greater
- *Solr 7.4.0*

Multi-host Services

For DeepNLP multi-host installations the following are automatically installed, as necessary:

- Hadoop
- Yarn 2.7.x or 2.8.x

Application Access Requirements

Although any service can be installed on any machine, consider the following:

- The DeepNLP web front end requires access to both `Postgres` and `Solr`
- The scheduler requires access to the web front end and `Solr`
- The scheduler must be able to access `Yarn` and `Hadoop`, when they are available